

# Bayesian model selection on linear mixed-effects models for comparisons between multiple treatments and a control

Lei Gong

Department of Statistics  
University of California, Riverside  
lei.gong@email.ucr.edu

Stephen R. Spindler

Department of Biochemistry  
University of California, Riverside  
spindler@ucr.edu

James M. Flegal

Department of Statistics  
University of California, Riverside  
jfflegal@ucr.edu

Patricia L. Mote

Department of Biochemistry  
University of California, Riverside  
mote@ucr.edu

September 28, 2015

## Abstract

We propose a novel Bayesian model selection technique on linear mixed-effects models to compare multiple treatments with a control. A fully Bayesian approach is implemented to estimate the marginal inclusion probabilities that provide a direct measure of the difference between treatments and the control, along with the model-averaged posterior distributions. Default priors are proposed for model selection incorporating domain knowledge and a component-wise Gibbs sampler is developed for efficient posterior computation. We demonstrate the proposed method based on simulated data and an experimental dataset from a longitudinal study of mouse lifespan and weight trajectories.

# 1 Introduction

Experiments are run by researchers in biology, medicine, and various other scientific fields, to compare multiple treatments with a control or standard treatment. Often these studies are conducted over a period of time and result in unbalanced repeated measured data that is commonly analyzed by the linear mixed-effects model (LMM). The LMM allows for some subsets of the regression parameters to vary among subjects, thereby accounting for sources of natural heterogeneity in the population. It models the mean response as a combination of population characteristics (fixed-effects) that are assumed to be shared by subjects, and subject-specific characteristics (random-effects) that are unique to a particular subject. It is common to introduce a set of fixed-effects for each group to model the effect of the treatment (see e.g. Fitzmaurice et al., 2004). To compare treatment groups with the control groups is, therefore, equivalent to comparing the sets of fixed-effects. Researchers are often interested in deciding which treatments are different from the control, and measuring the corresponding significance of the discrepancy.

Standard model selection procedures can be implemented to answer these questions (see e.g. Bolker et al., 2009; Fitzmaurice et al., 2004) with certain limitations. One can select models by using hypothesis tests (Stephens et al., 2005); that is, test simpler nested models against more complex models and report corresponding p-values. Although the likelihood ratio test (LRT) is widely used to determine the contribution of a factor in a model throughout statistics, it is not recommended by Pinheiro and Bates (2006) for testing fixed-effects in LMM, because of its unreliability for small to moderate sample size. Also, when the focus is to compare multiple treatments to a control, Burnham and Anderson (2002) criticize that such a pairwise comparison as an abuse of hypothesis testing. Another extensively used approach is the information-theoretic model selection procedure that allows comparison of multiple models (see e.g. Burnham and Anderson, 2002). This method relies on information criteria, such as Akaike information criterion and Bayesian information criterion (BIC), that use deviance as a measure of fit with a penalization on more complex models. Instead of reporting p-values, it estimates the magnitude of difference between models in expected predictive power and uses this to make a decision as to whether a variable should be included in the model or not. The resulting dichotomous decisions overly simplify the problem, and, in our case, withhold important information on the magnitude of the difference between a treatment and the control.

Motivated by these practical challenges faced by frequentist approaches, we resort to Bayesian model selection techniques (for a review see e.g. Clyde and George, 2004; George and McCulloch, 1997; Kuo and Mallick, 1998). In the Bayesian framework, this problem can be transformed to the form of parameter estimation (O’Hara et al., 2009). That is,

estimating the marginal posterior probability that a variable should be in the model, i.e. the marginal inclusion probability, which can usually be calculated directly from the posterior inference using an Markov chain Monte Carlo (MCMC) simulation.

There is an extensive literature on Bayesian model selection. George and McCulloch (1993); Geweke et al. (1996) develop the stochastic search variable selection (SSVS) technique for linear regression models that uses a Gibbs sampler to traverse the model space. Smith and Kohn (1996) extend its application to nonparametric regression models and show how integrating the regression parameters is essential to reliable convergence of a Gibbs sampler. Kohn et al. (2001) propose a more efficient single-site Metropolis-Hastings sampler. Holmes et al. (2002) consider selection and smoothing for a series of seemingly unrelated regressions. Chen and Dunson (2003); Kinney and Dunson (2007) develop variable selection for both fixed and random effects in generalized LMM. Recently, Bayesian model selection methods are extended to a series of spatially linked regression for functional magnetic resonance imaging analysis (see e.g. Lee et al., 2014; Smith and Fahrmeir, 2007). However, we are unaware of any work to extend Bayesian model selection on LMM to compare multiple treatments with a baseline.

In this article, we develop a novel Bayesian model selection approach on LMM that accommodates and compares multiple treatment effects. The method includes a re-parameterization of the fixed-effects of each treatment that attributes part of the effect to a baseline for direct measure of the difference between a treatment and the control. A modification of the fractional prior (Smith and Kohn, 1997) is proposed to undertake model selection and averaging, which is also related to Zellner’s g-prior (Zellner, 1986). The proposed prior incorporates information on subjects within the same group, which is critical to developing an efficient component-wise Gibbs sampler. This Bayesian paradigm provides practitioners with an intuitive understanding of the significance of each treatment through the marginal inclusion probability, which is inaccessible using existing techniques.

Our work is motivated by a longitudinal experiment of mouse lifespan and weight trajectories (see e.g. Spindler et al., 2014a,b, 2013a,b, 2014c) that aims to study how different treatments affect lifetime weight trajectories and identify potential longevity therapeutics. This application provides both a clear demonstration of our approach, and an example of enabling researchers to obtain previously unavailable information. However, the method itself is more general and applicable to most experiments that are interested in comparing multiple treatments to a control.

## 1.1 Experimental Data

The experimental data is from a longitudinal study of the lifespan and the weight trajectories of an F1 hybrid mouse (see e.g. Spindler et al., 2014a,b, 2013a,b, 2014c). The study is part of a compound screening program designed to identify potential longevity therapeutics, and it was approved by the Institutional Animal Care and Use Committee at the University of California, Riverside. It utilized an unbalanced statistical design to compare the lifespan and the weight trajectories of multiple treatment groups to that of one larger control group (Jeske et al., 2014). The disposition of dietary calories between body weight and metabolic energy appears to be a key to lifespan determination. In this article, we use a part of the dataset that recorded mouse body weight changes during the course of the experiment.

In the study, 2266 male C3B6F1 mice were initially fed a chow diet *ad libitum*. At 12 months of age (Day 365), 297 mice were shifted to daily feeding with 13.3 kcal/day/mouse of the control diet (Diet No.99), and the rest were shifted to control diet supplemented with one of 56 chemical, pharmaceutical, or nutraceuticals agents or combination of agents. All mice were fed daily and weighted bimonthly, but the number of mice progressively declined as the study progressed due to the onset of various age-related pathologies. The data are censored at extreme old age (Day 1369), when less than 1% of the mice remained.

The control and drug-treated mice gradually lost weight after the shift to the defined diets, which provided about 10% less than the *ad libitum* number of calories, to ensure the mice consumed all their food. Our main interest is to determine which supplemented diets significantly affected the lifetime weight trajectories. That is, researchers are interested in whether any deviation from the trajectory of the control group (Diet No.99) is statistically significant and is caused by dietary additions.

The rest of the paper is organized as follows. Section 2 formally introduces the Bayesian variable selection methodology. It outlines the re-parameterization of a LMM, prior specification, MCMC sampling schemes, and stopping criterion utilized. A simulation study is also detailed to evaluate the performance of the proposed method. Section 3 contains the empirical results from the analysis of the motivating example. Section 4 concludes with a discussion.

## 2 Model Selection on Linear Mixed-effects Models

In general, suppose that we have  $n$  subjects from  $G$  experimental groups under study, each with  $n_i$  observations taken repeatedly over time,  $i = 1, \dots, n$ , and let  $\mathbf{y}_i = (y_{i,1}, \dots, y_{i,n_i})^T$  denote the response vector for the  $i$ -th subject. Assume the  $i$ -th subject is from the  $g$ -th

group, for  $i = 1, \dots, n$ ,  $g = 1, \dots, G$ , let  $X_i$  and  $Z_i$  be two  $n_i \times p$  design matrices, then a LMM (Fitzmaurice et al., 2004; McCullagh and Nelder, 1989) is denoted as

$$\mathbf{y}_i = X_i \boldsymbol{\alpha}_g + Z_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N_{n_i}(\mathbf{0}, \sigma^2 I), \quad (1)$$

where  $\boldsymbol{\alpha}_g = (\alpha_{g,0}, \dots, \alpha_{g,p-1})^T$  are the fixed effects shared by subjects in the  $g$ -th experimental group. Further, denote  $\mathbf{b}_i = (b_{i,0}, \dots, b_{i,p-1})^T \sim N_p(\mathbf{0}, \lambda_D^{-1} I)$  as the random effects that are unique to the  $i$ -th subject, and hence we allow subject specific trajectories.

Note that, among the  $G$  groups, there is one control group and  $G - 1$  treatment groups. Without loss of generality, let us assume the  $G$ -th group is the control group, and  $g = 1, \dots, G - 1$  are the treatment groups. A primary goal for many experiments is to determine which alternative treatments significantly differ from the control group. To this end, we propose a re-parameterization of the fixed effects  $\boldsymbol{\alpha}_g$ 's in (1),  $g = 1, \dots, G$ . Let  $W_i$ ,  $X_i$  and  $Z_i$  be three  $n_i \times p$  design matrices, the re-parameterized model is denoted as, for  $i = 1, \dots, n$ ,  $g = 1, \dots, G$ ,

$$\mathbf{y}_i = W_i \boldsymbol{\alpha} + X_i \boldsymbol{\beta}_g + Z_i \mathbf{b}_i + \boldsymbol{\epsilon}_i, \quad \boldsymbol{\epsilon}_i \sim N_{n_i}(\mathbf{0}, \sigma^2 I), \quad (2)$$

where  $\mathbf{b}_i = (b_{i,0}, \dots, b_{i,p-1})^T \sim N_p(\mathbf{0}, \lambda_D^{-1} I)$  are the random effects as in (1), and  $\boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_{p-1})^T$  are the fixed effects of the control group,  $\boldsymbol{\beta}_g = (\beta_{g,0}, \dots, \beta_{g,p-1})^T$  are the fixed effects modeling the difference between the  $g$ -th group and the control group. That is, the group effect  $\boldsymbol{\alpha}_g$  in (1) is re-written as  $\boldsymbol{\alpha} + \boldsymbol{\beta}_g$  in (2), for  $g = 1, \dots, G$ . Also, it is straightforward to see, as the baseline,  $\boldsymbol{\beta}_G = (0, \dots, 0)^T$  for the control group.

Under the re-parameterization, the detection of significant treatments is equivalent to the identification of nonzero  $\boldsymbol{\beta}_g$ 's. To this end, we introduce 0/1 binary indicators  $\boldsymbol{\gamma}_g = (\gamma_{g,0}, \dots, \gamma_{g,p-1})^T$ ,  $g = 1, \dots, G$ , such that  $\beta_{g,j} = 0$  if  $\gamma_{g,j} = 0$  and  $\beta_{g,j} \neq 0$  if  $\gamma_{g,j} = 1$ . The  $\gamma_{g,j}$  is used to indicate whether the fixed effect on the  $j$ -th predictor of the  $g$ -th group differs from that fixed effect of the control group. Given  $\boldsymbol{\gamma}_g$ , let  $\boldsymbol{\beta}_g(\boldsymbol{\gamma}_g)$  be the vector of nonzero fixed effects and  $X_i(\boldsymbol{\gamma}_g)$  be the corresponding design matrix. Then, the model (2) can be written as, for  $i = 1, \dots, n$ ,  $g = 1, \dots, G$ ,

$$\mathbf{y}_i = W_i \boldsymbol{\alpha} + X_i(\boldsymbol{\gamma}_g) \boldsymbol{\beta}_g(\boldsymbol{\gamma}_g) + Z_i \mathbf{b}_i + \boldsymbol{\epsilon}_i. \quad (3)$$

This formulation allows us to look at the problem from the Bayesian SSVS perspective (George and McCulloch, 1993). The SSVS searches for models having high posterior probability by traversing the model space using MCMC techniques, and, thus, identifies subsets of predictors with nonzero coefficients. Moreover, it allows us to calculate the posterior distributions of the parameters by marginalizing over the other variables. In this way, the marginal inclusion probability can be obtained as a direct measure of the significance of

each treatment.

Note that, (3) is a very general setting that is applicable to a wide range of applications. It is possible to impose specific structures on  $\gamma$  to suit different scenarios to further simplify the modeling procedure. For example, given the setups of the motivating experiment, it is reasonable to assume a common intercept for all groups, since the mice were on the same diet at the first measurement; that is,  $\beta_{g,0} = 0$ , for  $g = 1, \dots, G$ . Therefore, given that the primary goal is to compare treatment groups to the baseline group  $G$ , it is desirable to impose the following settings on  $\gamma$ ,

$$\gamma_{1,0} = \dots = \gamma_{G-1,0} = \gamma_{G,0} = 0,$$

$$\boldsymbol{\gamma}_G = (\gamma_{G,0}, \dots, \gamma_{G,p-1})^T = (0, \dots, 0)^T.$$

## 2.1 Prior Specification

A proper prior must be placed on the nonzero coefficients  $\boldsymbol{\beta}_g(\boldsymbol{\gamma}_g)$  to undertake model averaging (see e.g. George and McCulloch, 1993; Kohn et al., 2001; Mitchell and Beauchamp, 1988; Smith and Kohn, 1996). In particular, Kohn et al. (2001); Smith and Fahrmeir (2007) propose a conditional prior for the coefficients by setting it proportional to a fraction of the likelihood. This fractional prior is related to the g-prior in Zellner (1986), and is located and scaled in line with the information from the likelihood. We propose a modification of this idea to accommodate multiple subjects within a group by setting  $\pi(\boldsymbol{\beta}_g(\boldsymbol{\gamma}_g)|y, \boldsymbol{\alpha}, \boldsymbol{\gamma}_g, b, \sigma^2) \propto \Pi_{i \in g} p(\mathbf{y}_i|\boldsymbol{\alpha}, \boldsymbol{\beta}_g(\boldsymbol{\gamma}_g), \boldsymbol{\gamma}_g, \mathbf{b}_i, \sigma^2)^{1/n_i}$ , so that

$$\boldsymbol{\beta}_g(\boldsymbol{\gamma}_g)|y, \boldsymbol{\alpha}, \boldsymbol{\gamma}_g, b, \sigma^2 \sim N\left(\hat{\boldsymbol{\beta}}_g(\boldsymbol{\gamma}_g), \sigma^2 \left(\sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g)\right)^{-1}\right), \quad (4)$$

where  $\hat{\boldsymbol{\beta}}_g(\boldsymbol{\gamma}_g) = \left(\sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g)\right)^{-1} \left(\sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) (\mathbf{y}_i - W_i \boldsymbol{\alpha} - Z_i \mathbf{b}_i)\right)$ , and  $\sum_{i \in g}$  stands for summation over all the subjects that belong to the  $g$ -th group.

This prior is proportional to the variance of the least squares estimate of  $\beta$ , and enjoys a number of attractive properties as pointed out by Kohn et al. (2001). The prior (4) is rescaled automatically if the design matrix  $X$  or the data  $y$  is rescaled because of its structure and the presence of  $\sigma^2$ . Moreover, this prior is invariant to location changes in  $\mathbf{X}$  and  $\mathbf{y}$  given the basis term  $(1, \dots, 1)^T$  is included in  $X$ . Also it is data-based since  $\hat{\boldsymbol{\beta}}_g(\boldsymbol{\gamma}_g)$  depends on  $y$ , which allows proper centering of  $\beta$ .

We consider the prior on  $\gamma$  to be  $\pi(\boldsymbol{\gamma}_g|\pi_g) = \prod_{j=0}^{p-1} \pi(\gamma_{g,j}|\pi_g)$ ,  $g = 1, \dots, G$ , where  $\pi(\gamma_{g,j}|\pi_g) \sim \text{Bernoulli}(\pi_g)$  and  $\boldsymbol{\pi} = (\pi_1, \dots, \pi_G)^T$  is a vector of hyper-parameters that

represents prior knowledge for every groups. Intuitively,  $\pi_g$  is the probability that researchers believe the  $g$ -th group is significantly different from the control group before conducting the experiment. For instance, we find a sensible setting, when there is little prior knowledge of the effects of the treatments, to be letting  $\pi_G = 0$  for the control group, and  $\pi_1 = \dots = \pi_{G-1} = 0.5$  for the  $G - 1$  treatment groups. We assume standard priors in Bayesian hierarchical models (see e.g. Gelman et al., 2004; Johnson and Jones, 2010; Smith and Kohn, 1996) for the rest of the parameters, i.e.  $\boldsymbol{\alpha}, b, \lambda_D, \sigma^2$ ,

$$\boldsymbol{\alpha} | \mathbf{d}_3, d_4 \sim N_p(\mathbf{d}_3, d_4^{-1})$$

$$\mathbf{b}_i | \lambda_D \sim N_p(\mathbf{0}, \lambda_D^{-1} I), i = 1, \dots, n$$

$$\lambda_D | d_1, d_2 \sim \Gamma(d_1, d_2)$$

$$\pi(\sigma^2) \propto 1/\sigma^2$$

where  $d_1, d_2, \mathbf{d}_3, d_4$  are hyper-parameters to be specified.

## 2.2 Posterior Inference

Combining the priors and likelihoods, the full joint posterior density for  $\theta = (\boldsymbol{\alpha}, \beta, \gamma, b, \sigma^2, \lambda_D)$  is characterized by

$$\begin{aligned} q(\boldsymbol{\alpha}, \beta, \gamma, b, \sigma^2, \lambda_D | y) &\propto \left[ \prod_{g=1}^G \left[ \prod_{i \in g} p(\mathbf{y}_i | \boldsymbol{\alpha}, \beta_g, \gamma_g, \mathbf{b}_i, \sigma^2) \pi(\mathbf{b}_i | \lambda_D) \right] \pi(\beta_g | \boldsymbol{\alpha}, \gamma_g, b, \sigma^2) \pi(\gamma_g) \right] \\ &\times \pi(\boldsymbol{\alpha}) \pi(\lambda_D) \pi(\sigma^2). \end{aligned} \tag{5}$$

This distribution has a complex form which we cannot sample from directly; instead, we resort to MCMC methodology for the posterior inference and employ a component-wise strategy (Johnson et al., 2013). Specifically, we propose a component-wise Gibbs sampler for posterior computation. To this end, we need the full conditional posterior distributions of each of the parameters in  $\theta$  to update the Markov chain. The derivation of the full conditional posterior distributions follows from (5) using straightforward algebraic route.

Schematically, we can set up a six-variable component-wise Gibbs sampler; that is, if we let  $\theta = (\gamma, \beta, \boldsymbol{\alpha}, \sigma^2, b, \lambda_D)$  be the current state and  $\theta' = (\gamma', \beta', \boldsymbol{\alpha}', (\sigma^2)', b', \lambda_D')$  be the future state, we iteratively sample from the full conditional posterior distributions to update the

chain,

$$\begin{aligned} (\gamma, \beta, \boldsymbol{\alpha}, \sigma^2, b, \lambda_D) &\rightarrow (\gamma', \beta, \boldsymbol{\alpha}, \sigma^2, b, \lambda_D) \rightarrow (\gamma', \beta', \boldsymbol{\alpha}, \sigma^2, b, \lambda_D) \rightarrow (\gamma', \beta', \boldsymbol{\alpha}', \sigma^2, b, \lambda_D) \\ &\rightarrow (\gamma', \beta', \boldsymbol{\alpha}', (\sigma^2)')', b, \lambda_D) \rightarrow (\gamma', \beta', \boldsymbol{\alpha}', (\sigma^2)')', b', \lambda_D) \rightarrow (\gamma', \beta', \boldsymbol{\alpha}', (\sigma^2)')', b', \lambda_D'). \end{aligned}$$

**Step 1.** The transition  $\gamma \rightarrow \gamma'$  consists of  $G \times p$  steps,

$$\begin{aligned} (\gamma_{1,0}, \gamma_{1,1}, \dots, \gamma_{1,p-1}, \dots, \gamma_{G,0}, \dots, \gamma_{G,p-1}) &\rightarrow (\gamma'_{1,0}, \gamma_{1,1}, \dots, \gamma_{1,p-1}, \dots, \gamma_{G,0}, \dots, \gamma_{G,p-1}) \\ &\rightarrow (\gamma'_{1,0}, \gamma'_{1,1}, \dots, \gamma_{1,p-1}, \dots, \gamma_{G,0}, \dots, \gamma_{G,p-1}) \\ &\vdots \\ &\rightarrow (\gamma'_{1,0}, \gamma'_{1,1}, \dots, \gamma'_{1,p-1}, \dots, \gamma'_{G,0}, \dots, \gamma'_{G,p-1}). \end{aligned}$$

From the Appendix, we have, for  $g = 1, \dots, G$  and  $j = 0, \dots, p-1$ ,

$$\begin{aligned} q(\gamma_{g,j} | \boldsymbol{\alpha}, \boldsymbol{\gamma}_{-(g,j)}, b, \sigma^2, y) &\propto \pi_g^{\gamma_{g,j}} (1 - \pi_g)^{1 - \gamma_{g,j}} \left( \frac{|\sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g)|}{|\sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g)|} \right)^{\frac{1}{2}} \\ &\times \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{i \in g} \boldsymbol{\phi}_i^T \boldsymbol{\phi} + \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right)^T \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right. \right. \\ &\quad \left. \left. - \left( \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right)^T \left( \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right] \right\}, \end{aligned} \tag{6}$$

where  $\boldsymbol{\gamma}_{-(g,j)} = (\gamma_{g,0}, \dots, \gamma_{g,j-1}, \gamma_{g,j+1}, \dots, \gamma_{g,p-1})^T$  and  $\boldsymbol{\phi}_i = \mathbf{y}_i - W_i \boldsymbol{\alpha} - Z_i \mathbf{b}_i$ .

At each step, an update is simulated from  $\gamma_{g,j}' \sim q(\gamma_{g,j} | \boldsymbol{\alpha}, \boldsymbol{\gamma}_{-(g,j)}, b, \sigma^2, y)$ . Since  $\gamma_{g,j}$  is binary, i.e.  $\gamma_{g,j} \in \{0, 1\}$ , the conditional posterior distribution  $q(\gamma_{g,j} | \boldsymbol{\alpha}, \boldsymbol{\gamma}_{-(g,j)}, b, \sigma^2, y)$  is easily normalized by evaluating (6) for  $\gamma_{g,j} = 0$  and  $\gamma_{g,j} = 1$ .

**Step 2.** The transition  $\beta \rightarrow \beta'$  consists of  $G$  steps,

$$\begin{aligned} (\beta_1, \beta_2, \dots, \beta_G) &\rightarrow (\beta'_1, \beta_2, \dots, \beta_G) \\ &\rightarrow (\beta'_1, \beta'_2, \dots, \beta_G) \\ &\vdots \\ &\rightarrow (\beta'_1, \beta'_2, \dots, \beta'_G). \end{aligned}$$

At each step, an update is simulated from a  $p$ -dimensional multivariate normal dis-



tribution,

$$\begin{aligned}\beta_g'(\gamma_g) &\sim q(\beta_g(\gamma_g)|\alpha, \gamma_g, b, \sigma^2, y) \\ &\sim N_{\sum_{j=0}^{p-1} \gamma_{g,j}} \left( V_1^{-1} \left[ \frac{1}{\sigma^2} \sum_{i \in g} X_i^T(\gamma_g)(\mathbf{y}_i - W_i \alpha - Z_i \mathbf{b}_i) \right], V_1^{-1} \right),\end{aligned}\quad (7)$$

where  $V_1 = \frac{1}{\sigma^2} \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\gamma_g) X_i(\gamma_g)$ .

**Step 3.** Consider updating  $\alpha$  where the update is simulated from a  $p$ -dimensional multivariate normal distribution,

$$\begin{aligned}\alpha' &\sim q(\alpha|\beta, \gamma, b, \sigma^2, y) \\ &\sim N_p \left( V_2^{-1} \left[ \frac{1}{\sigma^2} \sum_{g=1}^G \left( \sum_{i \in g} W_i^T \left( \mathbf{y}_i - X_i(\gamma_g) \beta_g(\gamma_g) - Z_i \mathbf{b}_i \right) \right. \right. \right. \\ &\quad \left. \left. \left. + \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\gamma_g) W_i \right)^T \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\gamma_g) X_i(\gamma_g) \right)^{-1} \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\gamma_g) (\mathbf{y}_i - X_i(\gamma_g) \beta_g(\gamma_g) - Z_i \mathbf{b}_i) \right) \right) \right] \right. \\ &\quad \left. + d_4 \mathbf{d}_3 \right], V_2^{-1} \right),\end{aligned}\quad (8)$$

where

$$V_2 = \frac{1}{\sigma^2} \sum_{g=1}^G \left[ \sum_{i \in g} W_i^T W_i + \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\gamma_g) W_i \right)^T \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\gamma_g) X_i(\gamma_g) \right)^{-1} \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\gamma_g) W_i \right) \right] + d_4.$$

**Step 4.** Consider updating  $\sigma^2$ . At each step, an update is simulated from a Inverse-Gamma

distribution,

$$\begin{aligned}
(\sigma^2)' &\sim q(\sigma^2 | \boldsymbol{\alpha}, \beta, \gamma, b, y) \\
&\sim \text{Inv-Gamma} \left( \frac{1}{2} (N + \sum_{g=1}^G \sum_{j=0}^{p-1} \gamma_{g,j}), \right. \\
&\quad \frac{1}{2} \sum_{g=1}^G \left[ \sum_{i \in g} \left( \mathbf{y}_i - W_i \boldsymbol{\alpha} - X_i(\boldsymbol{\gamma}_g) \boldsymbol{\beta}_g(\boldsymbol{\gamma}_g) - Z_i \mathbf{b}_i \right)^T \left( \mathbf{y}_i - W_i \boldsymbol{\alpha} - X_i(\boldsymbol{\gamma}_g) \boldsymbol{\beta}_g(\boldsymbol{\gamma}_g) - Z_i \mathbf{b}_i \right) \right. \\
&\quad \left. + \left[ \boldsymbol{\beta}_g - \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right]^T \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \right. \\
&\quad \left. \left[ \boldsymbol{\beta}_g - \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right] \right], \\
&\quad \left. \right)
\end{aligned} \tag{9}$$

where  $N = \sum_{g=1}^G \sum_{i \in g} n_i$ . Note that, we denote  $\text{Inv-Gamma}(\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} \exp\left(\frac{-\beta}{x}\right)$ , for  $x \in (0, \infty)$ , and  $\alpha, \beta > 0$ .

**Step 5.** The transition  $b \rightarrow b'$  consists of  $n$  steps,

$$\begin{aligned}
(\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_n) &\rightarrow (\mathbf{b}_1', \mathbf{b}_2, \dots, \mathbf{b}_n) \\
&\rightarrow (\mathbf{b}_1', \mathbf{b}_2', \dots, \mathbf{b}_n) \\
&\vdots \\
&\rightarrow (\mathbf{b}_1', \mathbf{b}_2', \dots, \mathbf{b}_n').
\end{aligned}$$

At each step, assuming the  $i$ -th subject is from the  $g$ -th group, an update is simulated from a  $p$ -dimensional multivariate normal distribution,

$$\begin{aligned}
\mathbf{b}_i' &\sim q(\mathbf{b}_i | \boldsymbol{\alpha}, \boldsymbol{\beta}_g, \boldsymbol{\gamma}_g, \sigma^2, \lambda_D, \mathbf{y}_i) \\
&\sim N_p \left( V_3^{-1} \frac{1}{\sigma^2} \left[ \frac{1}{n_i} Z_i^T X_i(\boldsymbol{\gamma}_g) \left( \sum_{j \in g} \frac{1}{n_j} X_j^T(\boldsymbol{\gamma}_g) X_j(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{\substack{j \in g \\ j \neq i}} \frac{1}{n_j} X_j^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_j + \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right. \right. \\
&\quad \left. \left. + Z_i^T \left( \mathbf{y}_i - W_i \boldsymbol{\alpha} - \left(1 + \frac{1}{n_i}\right) X_i(\boldsymbol{\gamma}_g) \boldsymbol{\beta}_g(\boldsymbol{\gamma}_g) \right) \right], V_3^{-1} \right),
\end{aligned} \tag{10}$$

where  $V_3 = \frac{1}{\sigma^2} Z_i^T Z_i + \lambda_D I + \frac{1}{\sigma^2} \frac{1}{n_i} Z_i^T X_i(\boldsymbol{\gamma}_g) \left( \sum_{j \in g} \frac{1}{n_j} X_j^T(\boldsymbol{\gamma}_g) X_j(\boldsymbol{\gamma}_g) \right)^{-1} X_i^T(\boldsymbol{\gamma}_g) Z_i$ .

**Step 6.** Finally, consider updating  $\lambda_D$ . At each step, an update is simulated from a Gamma distribution,

$$\begin{aligned}\lambda_D' &\sim q(\lambda_D|b) \\ &\sim \Gamma(\frac{np}{2} + d_1, \frac{1}{2} \sum_{g=1}^G \sum_{i \in g} \mathbf{b}_i^T \mathbf{b}_i + d_2).\end{aligned}\tag{11}$$

The posterior inference on model parameters can be carried out using the MCMC samples. Models with high posterior probability can be identified as those appearing most often in the MCMC output. One posterior quantity of interest is the marginal inclusion probability, i.e.  $\Pr(\gamma_{g,j} = 1|y)$ ,  $g = 1, \dots, G$ ,  $j = 0, \dots, p-1$ , which can be calculated using the proportion of draws in which  $\gamma_{g,j}$  is non-zero. It provides a direct measure of the significance of  $\beta_{g,j}$ , which remains challenging for the current frequentist methods. It, therefore, allows researchers for straightforward understanding of the significance of each treatment. Also, if needed, one may classify a treatment effect such that  $\Pr(\gamma_{g,j} = 1|y) > 0.8772$  as significant or otherwise insignificant (see e.g. Lee et al., 2014; Raftery et al., 1996; Smith and Fahrmeir, 2007).

### 2.3 Stopping Criterion

Determining how long to run an MCMC simulation is critical to performing legitimate posterior inference. Premature termination often runs the risk of getting inaccurate estimates. The relative standard deviation fixed-width stopping rule (FWSR) (see e.g. Flegal and Gong, 2015; Gong and Flegal, 2015) is implemented to terminate the MCMC simulation. It is a member of the FWSR family (for e.g. see Flegal and Gong, 2015; Flegal et al., 2008; Jones et al., 2006). The relative standard deviation FWSR is theoretically valid in that it terminates a simulation w.p. 1 and the resulting confidence interval achieves the nominal coverage probability. Moreover, it automates the stopping procedure for practitioners and outperforms convergence diagnostics in various numerical studies. Interested readers are directed to their papers for more details.

In short, the relative standard deviation FWSR terminates the simulation when the computational uncertainty is relatively small to the posterior uncertainty. Specifically, it controls the width of a confidence interval from a Markov chain central limit theorem through a threshold  $\epsilon$  and significant level  $\delta$ . Gong and Flegal (2015) also establish a connection between the standard deviation FWSR and using effective sample size (ESS) as a stopping criteria, i.e.  $K = 4z_{\delta/2}^2/\epsilon^2$ , where  $K$  is the number of effective samples and  $z_{\delta/2}$  is a critical value from the standard Normal distribution. Based on this connection,

for instance, setting  $\epsilon = 0.124$  and  $\delta = 0.05$  in the relative standard deviation FWSR is equivalent to terminate the simulation when an ESS reaches  $K = 1000$ .

## 2.4 Simulation Study

We report the results of a simulation study undertaken to validate the model and estimation procedure. The simulated dataset consists of a control group and five treatment groups. The control group is simulated based on estimates from maximum likelihood estimation (MLE) of a LMM on the control group of the experimental data. That is, denote  $Y_{i,t,99}$  as the weight of mouse  $i \in \{1, \dots, 297\}$  from the control group (Diet No.99) taken at time  $t \in \{365, 395, 456, 517, 578, 639, 700, 760, 821, 882, 943, 1004, 1065, 1125, 1186\}$ , we consider the following LMM based on (2),

$$Y_{i,t,99} = \alpha_0 + \alpha_1 t + b_{0,i} + b_{1,i} t + \epsilon_{i,t}, \quad \epsilon_{i,t} \sim N(0, \sigma^2), \quad (12)$$

where  $\alpha_0$  and  $\alpha_1$  are the global intercept and slope,  $b_{0,i}$  and  $b_{1,i}$  are the subject specific random effects, where  $\mathbf{b}_i = (b_{0,i}, b_{1,i})^T \sim N_2(\mathbf{0}, \lambda_D^{-1} I)$ , and  $\epsilon_{i,t}$  is the measurement error. Note that, as mentioned, the  $\beta$ 's in (2) are set to zero for the control group to serve as the baseline model.

Parameter estimation of (12) was carried out using the `lmer()` function in the R package `lme4` (Bates et al., 2012). Notice time was rescaled using  $t = (t - 365)/365$  prior to model fitting. The MLE estimates are  $\boldsymbol{\alpha} = (45.49, -5.75)^T$  and  $\sigma^2 = 5.06$ , and we set  $\lambda_D^{-1} = 1.0$ . Based on these estimates, we simulated 297 subjects from (12) as the control group.

We then simulated five treatment groups, each with 36 subjects, by adding  $\beta_{g,1}$ 's to (12), while keeping other settings the same as for the simulated control group,

$$Y_{i,t,g} = \alpha_0 + \alpha_1 t + \beta_{g,1} t + b_{0,i} + b_{1,i} t + \epsilon_{i,t}, \quad g = 1, \dots, 5, \quad (13)$$

where  $\beta_{g,1} \in \{-2.0, -0.5, 0.0, 0.5, 2.0\}$  for each group. To be consistent with the experimental settings, we artificially differentiate the slope of each treatment group by  $\alpha_1 + \beta_{g,1}$ , but maintained the same global intercept  $\alpha_0$ , since all mice were on the same diet at  $t = 0$ . Note that, we did not incorporate the “die-off” mechanism from the experiment into the simulation. Figure 1 shows the mean weight trajectories for this simulated dataset.

We followed the prior specification outlined in Section 2.1. The hyper-parameters  $d_1, d_2$  were set to  $d_1 = 0.001, d_2 = 0.001$  for the prior on  $\lambda_D$  to be vague. The hyper-parameters for  $\boldsymbol{\alpha} | \mathbf{d}_3, d_4$  were set using estimates obtained from a fitted LMM on the simulated control group. The prior inclusion probabilities for the treatment groups, i.e.  $\pi_g, g = 1, \dots, 5$ , were

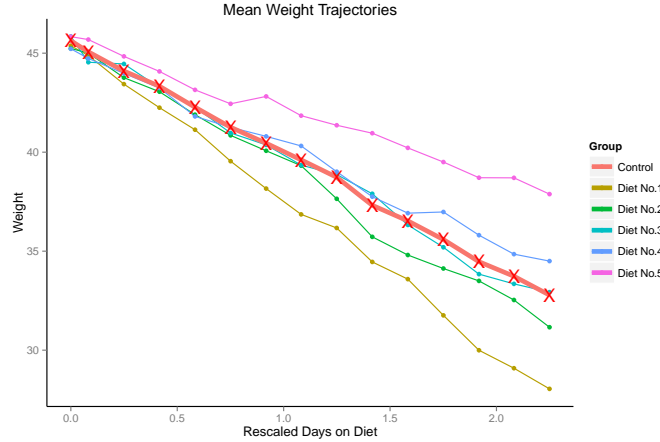


Figure 1: Mean weight trajectories for the 5 treatment groups and the control group.

set to 0.5 for equal probability between inclusion and exclusion.

The component-wise Gibbs sampler was run as described in Section 2.2. The simulation was terminated by the relative standard deviation FWSR with the tuning parameters  $\epsilon = 0.124$  and  $\delta = 0.05$ . It resulted in 16385 iterations with an effective sample size of at least 1000 for estimation of the posterior mean of all the parameters. The resulting MCMC outputs show the chain is mixing well and centered near the true parameter values.

Table 1: Fixed-effects estimates for the simulated dataset. For MLE, mean and 95% confidence interval (CI) are presented. For posterior inference, posterior mean, 95% credible interval (CI) and marginal inclusion probability (standard error in the parenthesis) are presented.

Parameter	Truth	MLE		Posterior		
		Mean	95% CI	Mean	95% CI	$\Pr(\gamma_{g,j} = 1 y)$
$\alpha_0$	45.50	45.570	(45.431, 45.570)	45.591	(45.484, 45.700)	
$\alpha_1$	-5.75	-5.708	(-5.852, -5.565)	-5.716	(-5.822, -5.612)	
$\beta_{1,1}$	-2.00	-2.130	(-2.546, -1.713)	-2.126	(-2.562, -1.685)	0.992(6.10e-5)
$\beta_{2,1}$	-0.50	-0.693	(-1.109, -0.276)	-0.698	(-1.126, -0.267)	0.983(7.44e-4)
$\beta_{3,1}$	0.00	-0.092	(-0.508, 0.325)	-0.093	(-0.518, 0.341)	0.442(3.88e-3)
$\beta_{4,1}$	0.50	0.708	(0.292, 1.125)	0.708	(0.283, 1.136)	0.987(5.74e-4)
$\beta_{5,1}$	2.00	2.266	(1.849, 2.683)	2.268	(1.830, 2.695)	0.992(6.10e-5)

We compare our results to the estimates from the frequentist approach (see e.g. Fitzmaurice et al., 2004), as it is widely used to model such problems. Researchers often combine LMM with certain model selection criteria, e.g. BIC, to determine which treatment are

significantly differ from the control (see e.g. Spindler et al., 2013a). Table 1 contains the posterior estimates from the proposed model, along with the MLE estimates from `lme4` for the LMM. Despite that both approaches result in estimates close to the truth, our approach introduces the marginal inclusion probability for each treatment group that is vital to straightforward interpretation and correct ranking of the significance of the treatment effects. If a dichotomous decision is desired, setting a threshold to the suggested value of 0.8772, one would correctly classify Diet No.3 to be insignificant relative to the control.

The sensitivity to the prior inclusion probabilities was also evaluated by repeating the simulation with  $\pi_g$ 's set to ranging from 0.3 to 0.7. We found no difference in model ranking, although the parameter estimates were slightly different. Other simulation settings showed comparable parameter estimations between our method and the MLEs, and correctness in model ranking, although the results are not shown here.

### 3 Application

In this section, we use the methodology detailed in Section 2 to analyze the mouse body weight data (see Section 1.1). Out of the 56 treatment groups in the original study, we limited our attention to 18 pre-screened treatments that the researchers are most interested in, as well as the control diet (Diet No.99). The 18 groups exhibited altered weight trajectories, or were related chemically to groups that did. Since all the groups consumed the same number of calories, weight trajectories are related to the disposition of dietary calories between body mass and metabolic energy, a key determinant of lifespan. For simplicity, we denote these 19 diets as  $\mathcal{G} = \{21, 22, 23, 24, 27, 28, 29, 34, 35, 39, 42, 43, 44, 45, 48, 53, 55, 63, 99\}$ . Similar to Section 2.4, the days on diet were rescaled prior to analysis. Figure 2a shows the mean weight trajectories for the 19 diet groups. Note that the mean weight estimates become unreliable as days on diet increases since mice died off in the process.

As previously pointed out, since the subjects were on the same diet at  $t = 0$ , it is reasonable to assume the same intercept for all groups. The individual weight trajectories suggest that, unlike the simulated dataset, a quadratic term is needed to characterize the trajectories. Specifically, we re-write the LMM from (3) as

$$Y_{i,t,g} = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \beta_{g,1}(\gamma_{g,1})t + \beta_{g,2}(\gamma_{g,2})t^2 + b_{0,i} + b_{1,i}t + b_{2,i}t^2 + \epsilon_{i,t}, \quad g \in \mathcal{G}. \quad (14)$$

Given the quantity of the data, it is possible to propose more complex models with additional polynomial terms. However, the additional terms complicate the model and the additional coefficients are difficult to interpret scientifically. Hence, we considered the model at (14) as our full model.

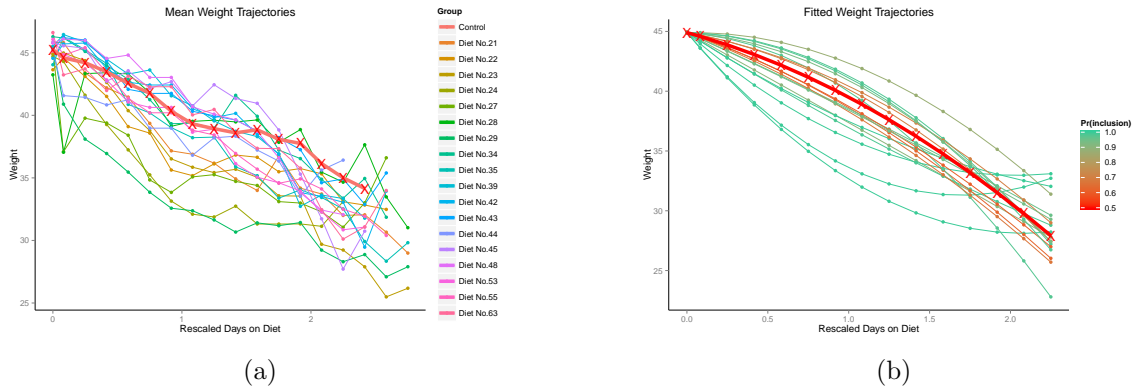


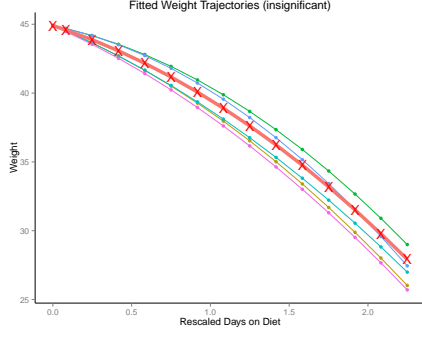
Figure 2: Weight trajectories for the 18 treatment diets and the control diet.

Priors were specified as in the simulation study and  $\gamma_{99}$  was set to  $\gamma_{99} = (0, \dots, 0)^T$  as the baseline model. The Gibbs sampler was terminated by the relative standard deviation FWSR with  $\epsilon = 0.124$  and  $\delta = 0.05$ , resulting in 115792 iterations with at least 1000 effective samples for estimation of posterior mean of the parameters related to fixed-effects and variance components. The resulting MCMC outputs show the chain is mixing well and indicate that the variation among subjects outweighs the variance of the measurement errors.

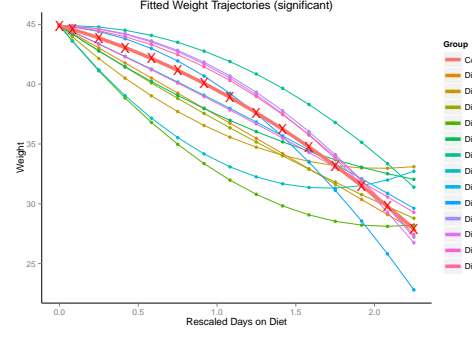
Table 2 contains the posterior estimates from the proposed model, along with the MLE estimates from the R package `lme4` for the LMM. The results from the proposed method are comparable to the MLEs in terms of point and interval estimates. However, the marginal inclusion probability from the proposed method provides a direct measure of the significance for each diet, which was unavailable in previous investigations using the frequentist approach.

Figure 2b shows the fitted weight trajectories colored based on the magnitude of their marginal inclusion probabilities. We can see that the treatment groups that have weight trajectories similar to the control group are the ones with lower inclusion probabilities. It shows that the marginal inclusion probability behaves well as a measure of the difference between a treatment diet and the control diet. Moreover, we find the suggested threshold 0.8772 is a reasonable value to classify treatment diets into significantly/insignificantly different from the control diet (see Figure 3a and Figure 3b).

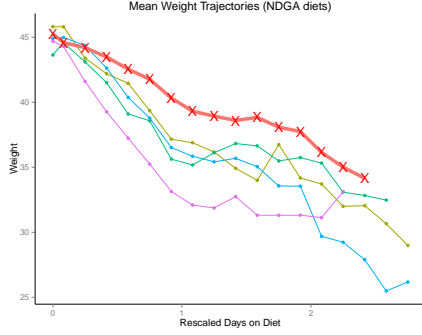
As an example, we compare four diets (Diet No.21 - Diet No.24) that are supplemented with 1.5, 2.5, 3.5, or 4.5 Nordihydroguaiaretic Acid (NDGA)/kg diet (Spindler et al., 2014c). Figure 3c shows the mean weight trajectories and Figure 3d shows the fitted weight trajectories for the 4 diets supplemented with NDGA and the control diet. We find that the fitted trajectories correctly capture the characteristics of the mean weight trajectories for each



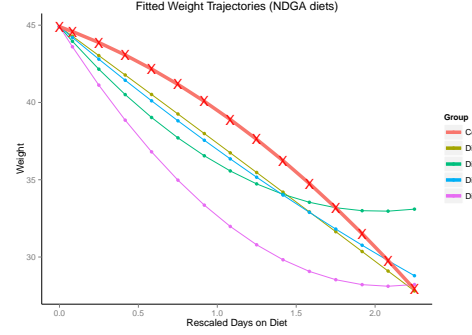
(a) Estimated weight trajectories for treatment diets classified into insignificant and the control diet.



(b) Estimated weight trajectories for treatment diets classified into significant and the control diet.



(c) Mean weight trajectories for 4 NDGA supplemented diets and the control diet.



(d) Estimated weight trajectories for 4 NDGA supplemented diets and the control diet.

Figure 3: Analysis of the experimental dataset based on the proposed model.

diet, especially for the first half of the experiment when most mice were alive. These actual and fitted weight trajectories indicate that NDGA produced a dose-responsive decrease in body weight in the absence of a change in food consumption. These data suggest NDGA may have extend mouse lifespan by decreasing calorie absorption, inducing a state of caloric restriction, or by increasing metabolic rate. Further experiments will be required to resolve these possibilities.

## 4 Discussion

This article proposes a novel method for Bayesian variable selection on LMM to compare multiple treatments with a control. It is built upon a modification of the fractional prior proposed by Smith and Kohn (1997) and a component-wise Gibbs sampler. It provides



practitioners with a framework to incorporate prior knowledge of each treatment, as well as an intuitive evaluation of its significance. This method is quite general and has a wide range of potential applications in fields such as biology and medicine.

The proposed method is advantageous in that multiple treatments are compared to a control group simultaneously. In addition, the Bayesian framework introduces marginal inclusion probabilities for each group that allow direct measure of the significance of each treatment, which is difficult using alternative frequentist approaches. Notice that, we introduce a vector  $\boldsymbol{\gamma}_g = (\gamma_{g,0}, \dots, \gamma_{g,p-1})^T$  in (3) as indicator instead of a single  $\gamma_g$  for each group, because it allows a more in-depth comparison between two groups. In this paper, the application on the experimental dataset provides new insights for researchers to group and study the diets based on their levels of significance.

We emphasize careful posterior inference when using MCMC methodology. One major challenge for practitioners is determining how long to run a simulation. While some simulations are so complex that a fixed time approach is the only practical one, this is not so for most experiments. We advocate the use of relative standard deviation FWSR (Flegal and Gong, 2015; Gong and Flegal, 2015), since it is proved to be easy to use, theoretically valid and superior to using convergence diagnostics as a stopping criteria (Flegal et al., 2008; Jones et al., 2006).

## Acknowledgements

The second author's work is partially supported by NSF grant DMS-13-08270.

## A Appendix

Full conditional posterior distributions are derived from (5), for  $i = 1, \dots, n$ ,  $g = 1, \dots, G$ , and  $j = 0, \dots, p-1$ . To calculate the full conditional posterior  $q(\gamma_{g,j} | \boldsymbol{\alpha}, \boldsymbol{\gamma}_{-(g,j)}, b, \sigma^2, y)$ , we integrate out  $\beta$  in (5) as Smith and Kohn (1996),

$$\begin{aligned}
q(\boldsymbol{\alpha}, \gamma, b, \sigma^2, \lambda_D | y) &= \int q(\boldsymbol{\alpha}, \beta, \gamma, b, \sigma^2, \lambda_D | y) d\beta \\
&\propto \left[ \prod_{g=1}^G \left[ \prod_{i \in g} \pi(\mathbf{b}_i | \lambda_D) \right] \pi(\boldsymbol{\gamma}_g) \right] \pi(\boldsymbol{\alpha}) \pi(\lambda_D) \pi(\sigma^2) \\
&\quad \times \prod_{g=1}^G \int_{\boldsymbol{\beta}_g} \prod_{i \in g} p(\mathbf{y}_i | \boldsymbol{\alpha}, \boldsymbol{\beta}_g, \boldsymbol{\gamma}_g, \mathbf{b}_i, \lambda_D, \sigma^2) \pi(\boldsymbol{\beta}_g | \boldsymbol{\alpha}, \boldsymbol{\gamma}_g, b, \sigma^2) d\boldsymbol{\beta}_g
\end{aligned} \tag{15}$$

To calculate (15), define  $\phi_i = \mathbf{y}_i - W_i \boldsymbol{\alpha} - Z_i \mathbf{b}_i$ . For a given  $g$ , consider

$$\begin{aligned}
& \int_{\boldsymbol{\beta}_g} \prod_{i \in g} p(\mathbf{y}_i | \boldsymbol{\alpha}, \boldsymbol{\beta}_g, \boldsymbol{\gamma}_g, \mathbf{b}_i, \lambda_D, \sigma^2) \pi(\boldsymbol{\beta}_g | \boldsymbol{\alpha}, \boldsymbol{\gamma}_g, \mathbf{b}, \sigma^2) d\boldsymbol{\beta}_g \\
& \propto \int_{\boldsymbol{\beta}_g} \left[ \prod_{i \in g} \sigma^{-n_i} \exp \left\{ -\frac{1}{2\sigma^2} \left( \boldsymbol{\phi}_i - X_i(\boldsymbol{\gamma}_g) \boldsymbol{\beta}_g(\boldsymbol{\gamma}_g) \right)^T \left( \boldsymbol{\phi}_i - X_i(\boldsymbol{\gamma}_g) \boldsymbol{\beta}_g(\boldsymbol{\gamma}_g) \right) \right\} \right] \\
& \quad \times \left| \frac{1}{\sigma^2} \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right|^{\frac{1}{2}} \times (2\pi)^{-\frac{1}{2} \sum_{j=0}^{p-1} \gamma_{g,j}} \\
& \quad \times \exp \left\{ -\frac{1}{2} \left[ \boldsymbol{\beta}_g(\boldsymbol{\gamma}_g) - \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right]^T \left( \frac{1}{\sigma^2} \sum_{i \in g} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right) \right. \\
& \quad \left. \left[ \boldsymbol{\beta}_g(\boldsymbol{\gamma}_g) - \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right] \right\} d\boldsymbol{\beta}_g \\
& = \sigma^{-\sum_{i \in g} n_i} \left( \frac{|\sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g)|}{|\sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g)|} \right)^{\frac{1}{2}} \\
& \quad \times \exp \left\{ -\frac{1}{2\sigma^2} \sum_{g=1}^G \left[ \sum_{i \in g} \boldsymbol{\phi}_i^T \boldsymbol{\phi}_i + \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right)^T \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right. \right. \\
& \quad \left. \left. - \left( \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right)^T \left( \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right] \right\}
\end{aligned}$$

Therefore, (15) is further simplified

$$\begin{aligned}
& = \lambda_D^{np/2} \exp \left\{ -\frac{\lambda_D}{2} \sum_{g=1}^G \sum_{i \in g} \mathbf{b}_i^T \mathbf{b}_i \right\} \prod_{g=1}^G \prod_{j=1}^p \pi_g^{\gamma_{g,j}} (1 - \pi_g)^{1 - \gamma_{g,j}} \exp \left\{ -\frac{1}{2} (\boldsymbol{\alpha} - \mathbf{d}_3)^T d_4 (\boldsymbol{\alpha} - \mathbf{d}_3) \right\} \\
& \quad \times \lambda_D^{d_1 - 1} \exp \{ -d_2 \lambda_D \} \prod_{g=1}^G \sigma^{-\sum_{i \in g} n_i} \left( \frac{|\sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g)|}{|\sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g)|} \right)^{\frac{1}{2}} \\
& \quad \times \exp \left\{ -\frac{1}{2\sigma^2} \sum_{g=1}^G \left[ \sum_{i \in g} \boldsymbol{\phi}_i^T \boldsymbol{\phi}_i + \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right)^T \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right. \right. \\
& \quad \left. \left. - \left( \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right)^T \left( \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right] \right\}
\end{aligned} \tag{16}$$

Based on (16), the full posterior distribution is characterized by

$$q(\gamma_{g,j} | \boldsymbol{\alpha}, \boldsymbol{\gamma}_{-(g,j)}, b, \sigma^2, y) \propto \pi_g^{\gamma_{g,j}} (1 - \pi_g)^{1 - \gamma_{g,j}} \left( \frac{|\sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g)|}{|\sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g)|} \right)^{\frac{1}{2}} \\ \times \exp \left\{ -\frac{1}{2\sigma^2} \left[ \sum_{i \in g} \boldsymbol{\phi}_i^T \boldsymbol{\phi}_i + \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right)^T \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{i \in g} \frac{1}{n_i} X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right. \right. \\ \left. \left. - \left( \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right)^T \left( \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) X_i(\boldsymbol{\gamma}_g) \right)^{-1} \left( \sum_{i \in g} (1 + \frac{1}{n_i}) X_i^T(\boldsymbol{\gamma}_g) \boldsymbol{\phi}_i \right) \right] \right\},$$

where  $\boldsymbol{\gamma}_{-(g,j)} = (\gamma_{g,0}, \dots, \gamma_{g,j-1}, \gamma_{g,j+1}, \dots, \gamma_{g,p-1})^T$ .

## References

- Bates, D., Maechler, M., and Bolker, B. (2012). lme4: Linear mixed-effects models using s4 classes.
- Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., and White, J.-S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3):127–135.
- Burnham, K. P. and Anderson, D. R. (2002). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer Science & Business Media.
- Chen, Z. and Dunson, D. B. (2003). Random effects selection in linear mixed models. *Biometrics*, 59(4):762–769.
- Clyde, M. and George, E. I. (2004). Model uncertainty. *Statistical Science*, pages 81–94.
- Fitzmaurice, G., Laird, N., and Ware, J. (2004). *Applied longitudinal analysis*. Wiley series in probability and statistics. Wiley-Interscience.
- Flegal, J. M. and Gong, L. (2015). Relative fixed-width stopping rules for Markov chain Monte Carlo simulations. *Statistica Sinica*, 25:655–676.
- Flegal, J. M., Haran, M., and Jones, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, 23:250–260.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, second edition.
- George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.
- George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica sinica*, 7(2):339–373.

- Geweke, J. et al. (1996). Variable selection and model comparison in regression. *Bayesian Statistics*, 5:609–620.
- Gong, L. and Flegal, J. M. (2015). A practical sequential stopping rule for high-dimensional Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, (just-accepted):00–00.
- Holmes, C., Denison, D., and Mallick, B. (2002). Bayesian model order determination and basis selection for seemingly unrelated regressions. *Journal of Computational and Graphical Statistics*, 11:533s551.
- Jeske, D. R., Flegal, J. M., and Spindler, S. R. (2014). Minimum size survival analysis sampling plans for comparing multiple treatment groups to a single control group. *Communications in Statistics-Theory and Methods*, 43(13):2689–2701.
- Johnson, A. A. and Jones, G. L. (2010). Gibbs sampling for a Bayesian hierarchical general linear model. *Electronic Journal of Statistics*, 4:313–333.
- Johnson, A. A., Jones, G. L., Neath, R. C., et al. (2013). Component-wise Markov chain Monte Carlo: Uniform and geometric ergodicity under mixing and composition. *Statistical Science*, 28(3):360–375.
- Jones, G. L., Haran, M., Caffo, B. S., and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101:1537–1547.
- Kinney, S. K. and Dunson, D. B. (2007). Fixed and random effects selection in linear and logistic models. *Biometrics*, 63(3):690–698.
- Kohn, R., Smith, M., and Chan, D. (2001). Nonparametric regression using linear combinations of basis functions. *Statistics and Computing*, 11(4):313–322.
- Kuo, L. and Mallick, B. (1998). Variable selection for regression models. *Sankhyā: The Indian Journal of Statistics, Series B*, pages 65–81.
- Lee, K.-J., Jones, G. L., Caffo, B. S., and Bassett, S. S. (2014). Spatial Bayesian variable selection models on functional magnetic resonance imaging time-series data. *Bayesian Analysis*, 9(3):699–732.
- McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, London, second edition.
- Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032.
- O’Hara, R. B., Sillanpää, M. J., et al. (2009). A review of Bayesian variable selection methods: what, how and which. *Bayesian Analysis*, 4(1):85–117.
- Pinheiro, J. and Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.

- Raftery, A. E. et al. (1996). Hypothesis testing and model selection via posterior simulation. *Markov chain Monte Carlo in practice*, pages 163–188.
- Smith, M. and Fahrmeir, L. (2007). Spatial Bayesian variable selection with application to functional magnetic resonance imaging. *Journal of the American Statistical Association*, 102(478):417–431.
- Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics*, 75(2):317–343.
- Smith, M. and Kohn, R. (1997). A Bayesian approach to nonparametric bivariate regression. *Journal of the American Statistical Association*, 92(440):1522–1535.
- Spindler, S. R., Mote, P. L., and Flegal, J. M. (2014a). Dietary supplementation with lovaza and krill oil shortens the life span of long-lived f1 mice. *Age*, 36(3):1345–1352.
- Spindler, S. R., Mote, P. L., and Flegal, J. M. (2014b). Lifespan effects of simple and complex nutraceutical combinations fed isocalorically to mice. *Age*, 36(2):705–718.
- Spindler, S. R., Mote, P. L., Flegal, J. M., and Teter, B. (2013a). Influence on longevity of blueberry, cinnamon, green and black tea, pomegranate, sesame, curcumin, morin, pycnogenol, quercetin, and taxifolin fed iso-calorically to long-lived, f1 hybrid mice. *Rejuvenation research*, 16(2):143–151.
- Spindler, S. R., Mote, P. L., Li, R., Dhahbi, J. M., Yamakawa, A., Flegal, J. M., Jeske, D. R., and Lublin, A. L. (2013b).  $\beta$ 1-adrenergic receptor blockade extends the life span of drosophila and long-lived mice. *Age*, 35(6):2099–2109.
- Spindler, S. R., Mote, P. L., Lublin, A. L., Flegal, J. M., Dhahbi, J. M., and Li, R. (2014c). Nordihydroguaiaretic acid extends the lifespan of drosophila and mice, increases mortality-related tumors and hemorrhagic diathesis, and alters energy homeostasis in mice. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, page glu190.
- Stephens, P. A., Buskirk, S. W., Hayward, G. D., and Martinez Del Rio, C. (2005). Information theory and hypothesis testing: a call for pluralism. *Journal of Applied Ecology*, 42(1):4–12.
- Zellner, A. (1986). On assessing prior distributions and bayesian regression analysis with g-prior distributions. *Bayesian inference and decision techniques: Essays in Honor of Bruno De Finetti*, 6:233–243.

Table 2: Fixed-effects estimates for the experimental dataset. For MLE, mean and 95% confidence interval (CI) are presented. For posterior inference, posterior mean, 95% credible interval (CI) and marginal inclusion probability (standard error in the parenthesis) are presented.

Parameter	MLE		Posterior		
	Mean	95% CI	Mean	95% CI	$\Pr(\gamma_{g,j} = 1 y)$
$\alpha_0$	44.879	(44.638, 45.120)	44.903	(44.720, 45.086)	
$\alpha_1$	-4.124	(-4.864, -3.384)	-3.687	(-4.134, -3.237)	
$\alpha_2$	-0.992	(-1.444, -0.540)	-1.718	(-2.076, -1.369)	
$\beta_{21,1}$	-3.671	(-6.637, -0.705)	-3.800	(-6.202, -1.432)	0.991(2.13e-4)
$\beta_{21,2}$	1.244	(-0.553, 3.041)	1.667	(-0.151, 3.475)	0.809(1.15e-3)
$\beta_{22,1}$	-7.894	(-10.828, -4.959)	-8.069	(-10.353, -5.789)	0.996(1.22e-5)
$\beta_{22,2}$	4.429	(2.644, 6.214)	4.611	(2.780, 6.461)	0.996(8.63e-6)
$\beta_{23,1}$	-4.467	(-7.282, -1.652)	-4.902	(-6.995, -2.810)	0.996(1.73e-5)
$\beta_{23,2}$	1.819	(0.203, 3.435)	2.352	(0.638, 4.046)	0.970(4.75e-4)
$\beta_{24,1}$	-12.373	(-15.317, -9.428)	-12.453	(-14.795, -10.103)	0.996(8.64e-6)
$\beta_{24,2}$	5.349	(3.533, 7.165)	5.596	(3.752, 7.444)	0.996(8.64e-6)
$\beta_{27,1}$	-6.643	(-8.694, -4.592)	-5.135	(-6.671, -3.570)	0.996(8.64e-6)
$\beta_{27,2}$	3.173	(1.948, 4.397)	3.098	(1.872, 4.344)	0.996(8.64e-6)
$\beta_{28,1}$	2.717	(0.585, 4.849)	3.899	(2.227, 5.611)	0.997(1.50e-5)
$\beta_{28,2}$	-0.969	(-2.273, 0.335)	-1.047	(-2.423, 0.316)	0.747(1.27e-3)
$\beta_{29,1}$	-13.462	(-15.574, -11.350)	-12.333	(-13.946, -10.704)	0.996(8.64e-6)
$\beta_{29,2}$	6.499	(5.195, 7.803)	6.432	(5.097, 7.751)	0.996(8.64e-6)
$\beta_{34,1}$	-0.768	(-2.898, 1.363)	-0.892	(-2.441, 0.638)	0.663(1.39e-3)
$\beta_{34,2}$	0.351	(-1.002, 1.704)	0.022	(-1.315, 1.459)	0.592(1.44e-3)
$\beta_{35,1}$	-1.552	(-3.607, 0.503)	-2.355	(-3.911, -0.752)	0.983(3.33e-4)
$\beta_{35,2}$	0.563	(-0.670, 1.796)	1.384	(0.140, 2.657)	0.920(7.83e-4)
$\beta_{39,1}$	3.183	(0.955, 5.410)	2.699	(0.829, 4.532)	0.975(4.29e-4)
$\beta_{39,2}$	-2.422	(-3.877, -0.967)	-2.212	(-3.707, -0.659)	0.976(4.16e-4)
$\beta_{42,1}$	3.625	(1.560, 5.689)	3.478	(1.874, 5.054)	0.996(2.86e-5)
$\beta_{42,2}$	-2.047	(-3.298, -0.795)	-1.686	(-2.953, -0.423)	0.963(5.28e-4)
$\beta_{43,1}$	1.627	(-0.442, 3.697)	1.298	(-0.201, 2.902)	0.815(1.13e-3)
$\beta_{43,2}$	-0.659	(-1.910, 0.591)	-0.364	(-1.705, 0.929)	0.584(1.45e-3)
$\beta_{44,1}$	-2.230	(-4.532, 0.068)	-1.028	(-2.777, 0.626)	0.675(1.37e-3)
$\beta_{44,2}$	0.853	(-0.626, 2.331)	0.274	(-1.195, 1.889)	0.549(1.46e-3)
$\beta_{45,1}$	1.094	(-1.122, 3.310)	1.357	(-0.351, 3.196)	0.772(1.23e-3)
$\beta_{45,2}$	-0.160	(-1.631, 1.310)	-0.696	(-2.308, 0.875)	0.633(1.41e-3)
$\beta_{48,1}$	4.118	(2.032, 6.204)	3.361	(1.695, 5.014)	0.996(6.29e-5)
$\beta_{48,2}$	-2.268	(-3.541, -0.995)	-1.728	(-3.020, -0.439)	0.960(5.48e-4)
$\beta_{53,1}$	-1.138	(-3.242, 0.966)	-1.381	(-2.955, 0.185)	0.813(1.14e-3)
$\beta_{53,2}$	0.047	(-1.343, 1.249)	0.176	(-1.181, 1.574)	0.530(1.47e-3)
$\beta_{55,1}$	-1.714	(-3.800, 0.372)	-2.442	(-4.072, -0.812)	0.983(3.27e-4)
$\beta_{55,2}$	0.695	(-0.584, 1.975)	1.356	(0.061, 2.658)	0.894(8.92e-4)
$\beta_{63,1}$	-2.483	(-4.590, -0.376)	2.712	(1.052, 4.376)	0.991(2.07e-4)
$\beta_{63,2}$	-1.367	(-2.678, -0.056)	-1.305	(-2.580, 0.019)	0.863(1.00e-3)